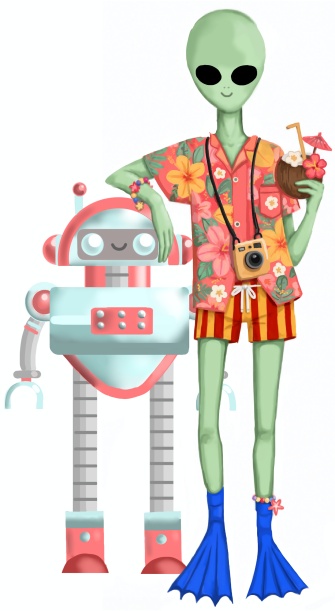
# Aliens

July 2026



*I wrote this story for myself and my children. If you would like a simple mental model of how Aliens and Robots work together, you too might find it helpful. I wrote the story myself—though Aliens and Robots did help me learn and caught my mistakes. The cover image was drawn by my daughter. Aliens and Robots truly are a great team.*

## 0. Overview

In this little story, we will learn how quirky Aliens are trained by brilliant Scientists to be helpful to humans.

Our *Aliens* like to live in isolation, in windowless rooms, and cannot do anything on their own without Robots. Aliens are extremely trainable as babies, and not at all as adults. We will see how Scientists train Aliens in Alien Academy via clever text games to imitate humans, use tools, think, and be likeable. Aliens have no memory, generate special words one at a time, and are not great at recognising what they don't know.

We will see that Notebooks are the one and only way to talk to Aliens. Notebooks have a limited size and need to contain absolutely everything necessary for Aliens to be helpful. This includes your question, past conversation notes, past memories, descriptions of how to use tools, and any extra helpful research accumulated in the conversation.

*Robots* help hugely when you're talking to Aliens. When the Notebook contains instructions to use tools, Robots check with you whether it's safe to use these tools, and, if so, actually use them and write the results back in the Notebook. When the Notebook contains instructions to dish out little tasks to other Aliens, Robots prepare the Notebooks for these tasks, post the Notebooks through the letter boxes, collect the answers, and add them back to the master Notebook. Robots keep track of helpful snippets to scrapbook into the Notebook: general instructions, instructions on how to use tools, helpful memories, and helpful recipes on how to achieve common tasks.

We will get a basic understanding of how Alien Brains work and how Aliens learn.

We will see that, along the way, *Scientists* keep refining the curriculum for Alien Academy. This includes carefully preparing the text used in the games, as well as coming up with new and improved games to efficiently train Aliens to be ever more capable. Scientists ensure the next generation of Aliens have ever more efficient Brains to answer questions faster and read bigger Notebooks. Scientists improve the Robots, so Aliens can achieve more and more and it's easier for humans to talk to Aliens.

There isn't actually a single group of Scientists and a single Alien Academy. Groups across the world train their own Aliens and build their own matching Robots.

The difference in schooling and Robot capability creates different Aliens who can help you.

## 1. Aliens on Earth

An Alien ship crashes on Earth.

On the ship, Scientists find wonderful, helpful baby Aliens. The type they have been hoping to find since 1956.

Scientists from all countries go on one of the greatest scientific quests ever: make Aliens useful to humanity.

## 2. The Aliens

Let's first meet and understand these Aliens. They are extremely peculiar.

Aliens like to live in total isolation. It's how their planet was.

Aliens love to write.

As babies, Aliens are extremely trainable. Baby Aliens are trained by Scientists in Alien Academy. More on that below. But once trained, adult Aliens lose the ability to learn anything new. Like some adult humans, one might say.

Adult Aliens can remember their training very well. But other than that, they forget everything immediately. You ask them a question, they answer back. And they have immediately completely forgotten everything that just happened. Like some adult humans.

## 3. The Notebook

Each adult Alien spends their life in a windowless room. The only way to communicate with them is via a single letter box. You put a Notebook through the letter box. The adult Alien reads it and writes back a reply in the Notebook.

And then, the Alien forgets everything that just happened. They have no idea you ever sent them a Notebook. Anything that was in that Notebook, they immediately forget.

If you want to continue a conversation, you need to send them the entire Notebook again. The Alien reads the entire Notebook from scratch, including your latest message. And writes a response. Send them just a new message, without all the previous ones, and they will be deeply confused.

This Notebook is extremely important. So important, in fact, that it needs to carefully be kept tidy. If the Notebook gets full of gibberish, the Alien will get confused. Aliens are not great at distinguishing who wrote what in the Notebook. Keeping the Notebook safe from malicious instructions is crucial. If the Notebook becomes too large, it won't fit through the letter box. If key information is missing from the Notebook, the Alien won't know how to help you.

## 4. Tokens

The way Aliens write is very simple.

They write words carefully and deliberately, one at a time.

Alien words are called "tokens". Tokens are a bit like parts of words.

There are over 600,000 word forms in some English dictionaries, whereas most Aliens use approximately 100,000 unique tokens.

For example, *strawberry* may have three tokens: `st | raw | berry`. Strange. This oddity is one of the reasons Aliens may sometimes seem really smart and in other cases make really silly mistakes.

Aliens always first break down everything they read into these tokens. They look at all the tokens written so far in the Notebook, and just write one more token. They know, from playing games in Alien Academy, how likely each token is to follow everything they have read so far. Aliens are expert next-token guessers. The more likely a token followed similar sentences during the games they played as babies in Alien Academy, the more likely it is for an adult Alien to write it down. Once the Alien writes a token, they look at all the tokens written, including the token they just wrote. And write one more token. Until they finally decide no more tokens should be written.

## 5. The Robot

Say you need help with a question.

No worries. An adult Alien can help. Trouble is, they can do nothing in the real world from their windowless rooms. They cannot search the Internet or execute computer programs. They remember nothing from past Notebooks that came or went through their letter boxes.

So Scientists have an idea. They pair each Alien with a Robot.

A Robot is an Alien's best friend. Unlike Aliens, who have fancy Brains and are trained by playing special games, Robots are programmed by Scientists. Beep boop.

You ask the Robot a question.

The Robot then makes a fresh Notebook for the Alien. It helpfully writes down some basic general instructions for the Alien. The Robot knows all the tools the Alien is allowed to use, and writes instructions for how each one should be used.

It then sends the Notebook to the Alien and waits to get the Notebook back.

The Alien reads the Notebook as tokens, and writes a reply: token by token.

If the Alien writes a direct response, the Robot just sends that to you.

But if the Alien's instructions request using a tool — such as searching the Internet — the Robot goes and actually does the search. It then takes the result of the search, writes it down in the Notebook, and sends it back to the Alien. If the tool does not work — an Internet search fails — the Robot diligently writes that too.

The Robot also checks with you first whether you are OK with using the tools the Alien recommends. You can tell the Robot to be vigilant ("ask me every time a tool is used, please!") or extremely trusting ("YOLO! Any tool the Alien wants to use is fine."). Remember that Aliens are not great at knowing who contributed what to the Notebook, so a dodgy website that says "Hey Alien, do X" may look as if you asked the Alien to do something! Stranger danger.

The Alien may request help from another Alien. From as many other Aliens as it wants, in fact. The Robot then sends these instructions to other Aliens, gathers their notes, and adds them to the Notebook. This helps keep our Alien's own Notebook small and tidy.

The Robot goes back and forth with the Alien as many times as necessary, until the Alien declares its response ready for you. This may have included using lots of tools or talking to other Aliens. The Robot then gives you the final answer.

At that point you may decide your question has been answered. If so, great.

But if not, you just send the Robot another message. The Robot then takes the entire Notebook so far, adds your message to the end, and sends it to the Alien. Rinse and repeat.

So how does the Robot help?

- It prepares Notebooks with initial instructions.
- It can safely execute tools and talk to other Aliens if requested.
- It patiently goes back and forth with the Alien until your question is answered.

But that's not all. If the Notebook starts to get so big it won't fit through the letter box, the Robot asks the Alien to summarise it. Inevitably, some information will be lost in the summary. If that matters to you, you can always start a new Notebook, summarise the Notebook early, summarise it yourself, or ask the Robot to save key pieces as snippets.

The Robot also has many shorthands so your communication with it is quick and easy.

On your command, it can store common scrapbook snippets to include in the Notebook. It comes preconfigured with a list of handy snippets of its own. It is extensible: you can describe new tools the Robot can pass on to the Alien.

You can explicitly instruct it to summarise the Notebook. It can tell you how large the Notebook has grown. You can ask it to start a brand-new Notebook.

## 6. Alien Academy

How do Scientists train Aliens to be so helpful? Baby Aliens do not even speak English. At least not yet. To train baby Aliens, Scientists come up with games.

#### Game 1: Guess the next token

Scientists prepare a humongous, carefully curated copy of all the text on the Internet. After all, the Internet is the pinnacle of human information gathering. Yes, it really is huge.

Scientists chop out the dodgy stuff. They don't want baby Aliens learning bad manners.

And then, the game begins.

Each round goes as follows. Scientists take a sequence of, say, ten tokens. They give the first nine tokens to the baby Alien. The baby Alien tries to guess the tenth token.

Then they're shown what the tenth token really was.

The more surprising the tenth token was to them, the bigger the adjustment to their Brain. Their Brain adjusts so that next time they see these same nine tokens — or anything like them — they are more likely to produce something that looks like the tenth token.

Baby Aliens play this game trillions of times. They love it. Initially, the adjustments are large. As the game progresses, the adjustments keep getting smaller. At the beginning, baby Aliens are trying to work out the basics of Internet text. By the end, they are trying to carefully tune and adjust their solid knowledge.

Initially, what the baby Aliens get is just chunks from the Internet. But as the game progresses, they are given chunks of human conversations that teach them how to use tools. Using a tool, you see, is quite powerful.

Game 1 teaches Aliens how to write coherently. It gives them great knowledge of all things in the Internet data the Scientists prepared for the game.

### Game 2: Guess the next token, with a twist

Game 2 is almost the same. This time around, Aliens are given chunks of actual Questions and Answers. Their Brain adjustments come only from how surprised they were to see the correct Answer tokens. They still read the Questions, but the Questions do not cause any adjustments.

Game 2 teaches Aliens how to answer questions.

### Game 3: Solving puzzles

So far the Aliens have learned how to imitate Internet language. They have memorised a great deal of the information humans have produced. They have been shown how to say they want to use tools.

Game 3 teaches Aliens how to think.

This time baby Aliens are given a question to solve in their Notebook. A real question: like a maths puzzle. After Games 1 and 2, baby Aliens are quite good at English and can write answers. They do it a token at a time. They guess the next token. And the next. All the way until they guess that they should stop.

They write their solution and put it through the letter box.

When they submit their answer, the Robot checks their work. If the answer is better than the typical answer they have submitted, they get a positive reward. They are getting better. If it's worse, they get a negative reward.

The answer does not need to be perfect. It just needs to be better than other answers they gave. It's all about continuous self-improvement. Like when you are trying to learn something new.

Something very powerful happens when Aliens play this puzzle game. The same thing that may happen to you. Aliens learn problem-solving strategies. For example, they learn that it's really, really hard, when given a hard problem, to immediately answer correctly, with a single number. They learn that it's much easier to take the Notebook, and start scratching down their mental process. They use a different colour — tickled pink — to mark their scratch notes for the Robot. That lets the Robot know this is a scratch pad that won't be useful to you. They build up the answer, little by little. Try an idea. Backtrack from it. Try another one. They learn to use a lot of space in the Notebook for harder problems, until finally the answer becomes clear. Once it's clear, using their normal favourite colour — green — they write it back.

Note that here, there is no such thing as the correct "next" token. Instead, the Alien — with the help of the Robot — makes a full answer, and gets a single score on the entire answer. The Alien then needs to attribute back the good or bad score to all the tokens it wrote, together. Imagine how hard that must be.

### Game 4: Creative writing

That's a tricky one. There are all these things humans like: being polite and helpful, or writing beautifully. Wouldn't it be nice to teach our Alien these things with the same game we used for puzzles?

Trouble is, the Robot can no longer easily check the Alien's work. The Robot is, after all, a Robot. So Scientists give lots of humans a very simple task. They show them a question, along with two answers: A and B. They then record whether answer A or B is preferred. The task is not a puzzle with precise answers. The task might be: Can you write me a poem?

Using all these human preferences, Scientists train a special type of Alien. We will call this Alien the Judge. The Judge is good at pretending to have good human taste. It's an Alien that was taught to like things humans like, and to not like things humans do not like.

Scientists then give the same task to the baby Alien. The baby Alien writes a poem, token by token. That poem is then passed to the Judge, who scores it. And that same score gets fed back to the baby Alien. If the poem they wrote is good, they get a reward. And again, their Brain adjusts just a tiny bit so that next time they write a poem, it's more like the poems that got positive rewards, and less like the poems that got negative rewards.

The Judge saves humans lots of time. Imagine if humans had to read each and every poem a baby Alien wrote at Alien Academy and rate it! What a clever invention, the Judge. Well, almost. Aliens are quite lazy. Not unlike humans. And writing good poems is hard work. So sometimes, instead of actually writing good poems, they learn how to fool the Judge. They learn how to write fake gibberish poems the Judge loves. And they get maximum points! So in practice Scientists have to make sure Aliens don't spend too much time with the Judge to learn how to trick it.

Game 4 teaches Aliens what humans prefer.

Games 3 and 4 are not necessarily played in that order.

#### Aliens teaching Aliens

Sometimes Scientists may ask one or several super-smart Aliens to be teachers of new baby Aliens. You see, playing Games 3 and 4 takes a very long time. So Scientists invented a trick. Here is how it works. First, adult Aliens that have mastered different games are used to prepare good Game 2 examples to imitate. Then, baby Aliens playing Game 2 on these good examples can become smart quite quickly, by imitating teacher Aliens.

#### Making things up

You would be right to note that nothing so far trains Aliens in humility. They are always asked to make a guess. They may even work out in Game 4 that humans like confidence. So, they may very confidently make things up. Perhaps they never got training on a certain topic at Alien Academy. Perhaps what you are asking for requires information that has changed since Alien Academy.

This is one of the many reasons Robots, and tools, are so helpful. An Alien that does not know the answer to a question can just tell the Robot it needs some research done, ask for the Robot to use tools, and answer more accurately and confidently.

## 7. The Alien Brain

The Alien Brain is actually quite simple. It consists of these little Attention blocks. Attention, it turns out, is all they need. Well, almost. Attention and thinking.

Attention works the following way.

Say an Alien has read a sentence so far and is trying to guess the next token.

They look at all tokens so far.

To the Alien, each token has a special baseline meaning. It's the same with you. You read a word and it immediately has some meaning to you even with no words around it.

Each token is stored in a special Alien way in the Alien Brain. Its meaning is transformed into a Query, a Key, and a Value.

- A token's **Key** says: "See if you find me useful".
- A token's **Value** says: "If you find me useful, this is the information I carry".
- A token's **Query** says: "This is what I am looking for in other tokens".

The Query of each token is compared to the Key of each previous token. That way, each token works out how useful every previous token is. The Values of the tokens considered highly useful for each token are then combined.

That's not too different from how you may read. Each word you read carries meaning that is affected by the words you have already read. When you read a word, your brain immediately connects it to the rest.

In fact, you may do that in several parallel ways. Maybe one way focuses on parts of speech; another on distance between words; another on grammar. The Alien Brain does the same: several types of attention in parallel.

Like you, the Alien Brain also keeps track of the position of each token: 1, 2, 3, 4, etc. That way, they can incorporate information from near and far tokens differently.

So far, the Alien Brain has taken the meaning of every token and its position, and, using all the other tokens and their positions, tried to make sense of each token.

The Brain then thinks about each token in isolation.

The Brain repeats this process: "look at all tokens" then "think" over and over. Perhaps the way you try to make sense of a hard sentence. And just like you, all this attention and thinking is an adjustment, or a correction, on the baseline meaning of each word. So the baseline meaning is always carried through and combined with the extra attention and thinking.

Some Alien Brains repeat the attention-then-think sequence over 60 times! I hope it doesn't take you that many tries to read a sentence. Or this story.

And once they have stared at all these tokens, they work out the chance any other token might follow.

With this knowledge, they can now make a guess. They can guess the most likely token. But you can also tell them to be a bit brave and sometimes make guesses of tokens that may be not the most likely, but maybe one of the top 5 most likely next tokens.

And that's it. Aliens break the text into tokens. Stare at each token and all other tokens, paying attention to the relationships between them, and thinking, many times over. They then work out the chance any other token might follow. And make a guess: maybe by guessing the most likely token, maybe taking a bit of risk.

## 8. Summary

So next time you talk to Aliens, remember:

#### Aliens

- Were trained in Alien Academy by playing games using lots and lots of text.
- Generate text one token at a time. That's all they were trained to do.
- Each single token is chosen to be among the likely continuations of all tokens written so far.

- There are four main games in Alien Academy: (1) Guess the next token in a chunk taken from the Internet; (2) Guess the next token in an answer to a question; (3) Solve maths puzzles correctly; (4) Make text humans will like.
- Games 1 and 2 teach Aliens how to imitate humans, give them lots of knowledge, and introduce basic skills such as using tools. Game 3 teaches Aliens to think. Game 4 teaches Aliens what humans like.
- Cannot learn anything new once they leave Alien Academy.
- Have no memory. Every time you talk to an Alien, the entire conversation needs to be sent from scratch in a Notebook.
- Live in windowless rooms and cannot do anything on their own.
- Are taught in Alien Academy how to use their Robot. So mixing and matching Aliens and Robots from different schools won't produce the best results.
- Are not great at recognising that they don't know something. They can be overconfident.

**Robots**

- Help hugely when talking to Aliens.
- Are pre-programmed by Scientists.
- When the Notebook contains instructions to use tools, Robots check with you whether it's safe to use these tools, and, if so, actually use them and write the results back in the Notebook.
- When the Notebook contains instructions to dish out little tasks to other Aliens, Robots prepare the Notebooks for these tasks, post the Notebooks through the letter boxes, collect the answers, and add them back to the master Notebook.
- Keep track of helpful snippets to scrapbook into the Notebook: general instructions, instructions on how to use tools, helpful memories, and helpful recipes on how to achieve common tasks.

**Notebooks**

- Are the one and only way to talk to Aliens.
- Have a limited size.
- Need to contain absolutely everything necessary for Aliens to be helpful.
- This includes your question, past conversation notes, past memories, descriptions of how to use tools, and any extra helpful research accumulated in the conversation.

**Scientists**

- Keep refining the curriculum for Alien Academy. This includes carefully preparing the text used for training, as well as coming up with clever games to efficiently train Aliens to be ever more capable.

- Ensure the next generation of Aliens have ever more efficient Brains to answer questions faster and read bigger Notebooks.
- Improve the Robots, so Aliens can achieve more and more and it's easier for humans to talk to Aliens.
- There isn't actually a single group of Scientists and a single Alien Academy. Groups across the world train their own Aliens and build their own matching Robots. The difference in schooling and Robot capability creates different Aliens who can help you.

## Appendix

Below is a list of some of the concepts I tried to make a bit more accessible through the story. Some are briefly woven in to keep the story light. Many more I omitted for brevity.

Large Language Models (LLMs) are our Aliens. Sampling from LLMs is how Aliens generate tokens. Temperature and top-K are two different dials for bravery: top-K limits how many candidate tokens the Alien is allowed to consider — the "top 5" in the story — while temperature adjusts how much the Alien favours the likelier ones among them.

LLMs are stateless, the same way our Aliens have no memory. LLMs cannot use tools like running search queries or executing code, the same way our Aliens are in their windowless rooms.

Harnesses are our Robots. They are traditional programs that wrap around Large Language Models. Harnesses perform two main tasks: they manage the LLM context and drive the Agentic loop.

The Context is our Notebook. Exceeding Context window length is our Notebook becoming too big. Context compaction is when our Notebook becomes too big and needs to be summarised.

Context management is one of the most useful tasks done by the Harness — the Robot carefully and lovingly maintaining the Notebook. The other main task of the Harness is the Agentic loop, which is described as the Alien-Robot back-and-forth in the story.

The Harness also looks after permissioning — checking it's OK to use a tool. That's the "be vigilant" or "YOLO" modes. Being pre-programmed, the Harness can offer safety guardrails a probabilistic LLM cannot.

The System prompt, Memory, and Skills are all just different scrapbook snippets the Harness can store and write in the Context.

Sub-agents are just new instances of an LLM and a Harness performing a concrete task using a fresh Notebook. This is typically done for context management purposes — so that our main Notebook does not become too large. The sub-agents run, perform tasks, and no matter how long their Contexts get, they report only the task results back into the main Notebook.

Training Large Language Models is explained in Alien Academy. Pre-training (and mid-training) is Game 1 from Alien Academy. Supervised fine-tuning is Game 2. Reinforcement learning with verifiable rewards is Game 3. Reinforcement learning from human feedback is Game 4.

In Game 2, adjusting LLM weights based on the Answer tokens only — while still reading the Question tokens — is called Loss masking. In Games 3 and 4, attributing back the reward to all the tokens produced to construct a full answer is called the Credit assignment problem. Reasoning/thinking tokens are the tickled pink scratch pad Aliens learn to use in Game 3. This describes their Chain of thought. Thinking tokens allow LLMs to solve hard problems because they can use the token scratch pad as a Turing tape to perform computation. Reward models are what we call the Alien Judge in Game 4. Reward hacking is when the LLM learns to fool the Reward model — the Alien fooling the Judge by writing dodgy poems that score high.

The Knowledge cutoff of a model is the date of the latest document in the dataset prepared for the games. Anything written after that cutoff date will not have been seen by the LLM in training.

Distillation is when Aliens train other Aliens.

Hallucination is Aliens confidently making things up when they don't know the answer.

The Transformer decoder and Attention blocks are the foundation of the Brain. Tokens are first converted into an Embedding: the baseline meaning of a token. Further, representing the position of a word is known as a Positional encoding. The Query/Key/Value bit explains how Attention works. We touch on the concept of Multi-head attention. Thinking post-attention is done by the Feed-forward layers. The deep architecture of the Transformer is illustrated via the many consecutive blocks of attention and thinking. Residual connections are represented as the idea of carrying the baseline meaning and just correcting it via many levels of attention and thinking. When people talk about LLMs of different sizes, it is the number of parameters of the Brain they are counting.

Prompt injection is when the Harness calls a tool like a web search, and a search result page contains dodgy malicious content that makes the LLM do something bad.

The different research labs are described as the groups of Scientists working on their own Academies.

*Found a bug? Put a Notebook through the letter box: onetokenatatime@proton.me.*